Annual sums of carbon dioxide exchange over a heterogeneous urban landscape through machine learning based gap-filling

Menzer, O. et. Al., *Atmospheric Environment* 101 (2015) 312-327
doi:10.1016/jatmosenv.2014.11.006

Objectives:

- This paper presents a novel wind-direction-dependent, machine-learning, gap-filling model approach to calculate annual sums of CO_2 flux for three years of eddy covariance (EC) observations in a suburban landscape.
- The gap-filling modeling framework uses machine learning to select explanatory variables and temporal variables, and then constrains models separately for spatially classified subsets of the data.
- Total net CO_2 fluxes (F_C) over a suburban neighborhood of Minneapolis-Saint Paul, Minnesota, USA, were measured from June 27, 2006 to June 26, 2009 at 40 m on the tall tower using an EC system consisting of a 3-D sonic anemometer and an infrared gas analyzer.
- Thirteen variables, including meteorological observations, traffic counts and satellite-derived greenness indices were selected as potential explanatory variables for the gap-filling models; six variables were chosen to be used in the final models through preliminary analysis.
- Three different methods (Artificial Neural Network (ANN), Radial Basis Function Network (RBF), and Gaussian Process (GP)) were used for model training.
- The study attempted to identify the controls of CO_2 emission specific to the urban environment and how they varied in time and space (throughout the tower footprint).
- It also trained machine learning regression models that could reproduce fluxes measured in spatially heterogenous landscapes, and then evaluated model performance and error.
- The study also assessed the heterogeneity of the flux source area using various potential explanatory variables and determined their importance for modeling and gap-filling of EC measurements in the urban environment.
- The seasonal and annual CO_2 flux sums for the tower site observed footprint and for the main use types within it were calculated, and associated uncertainties were estimated.

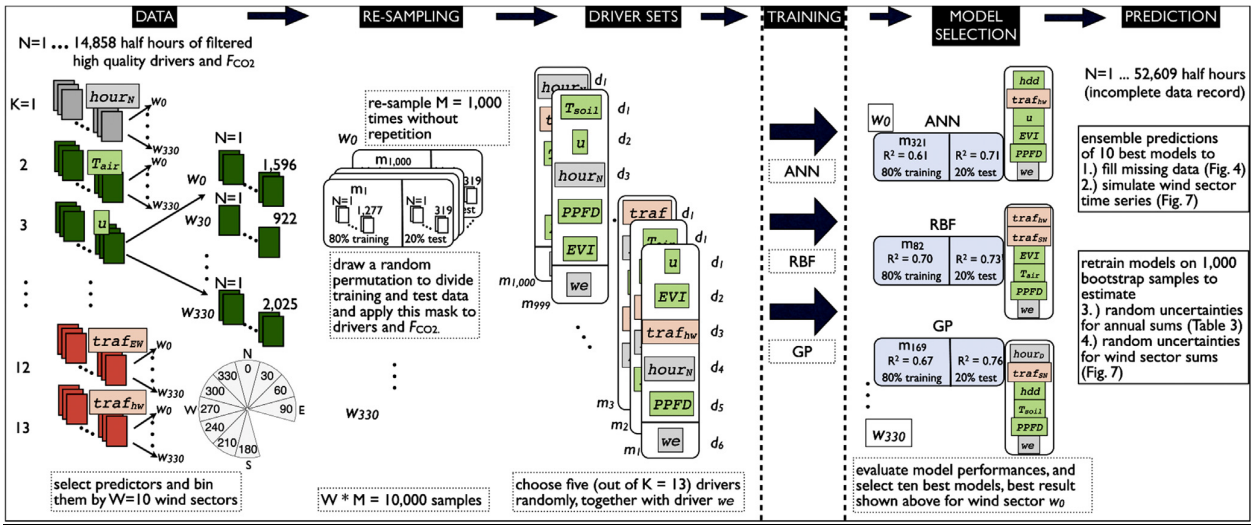
New Science:

- Monthly carbon budgets simulated by the gap-filling models were in good agreement with an ecophysiological bottom-up study at the same site explaining 64-88% of the variability in the fluxes.
- Total annual carbon dioxide flux sums for the tower site ranged from 1064 to 1382 g C m^{-2} yr^{-1}, across different years and different gap-filling methods.
- Bias errors of annual sums resulting from gap-filling did not exceed 18 g C m^{-2} yr^{-1} (1.8% of the annual flux) and random uncertainties did not exceed ±44 g C m^{-2} yr^{-1} (±3.8% of the annual flux).
- Regardless of the gap-filling method used, the year-to-year differences in carbon exchange at this site were small.
- Modeled annual wind sector F_C budgets, calculated by predicting artificial time series with models that had been trained for each 30 degree wind sector separately, differed by a factor of two depending on wind direction.
- The above results indicated that the modeled time series captured the spatial variability in both the biogenic and anthropogenic CO_2 sources and sinks in a reproducible way.

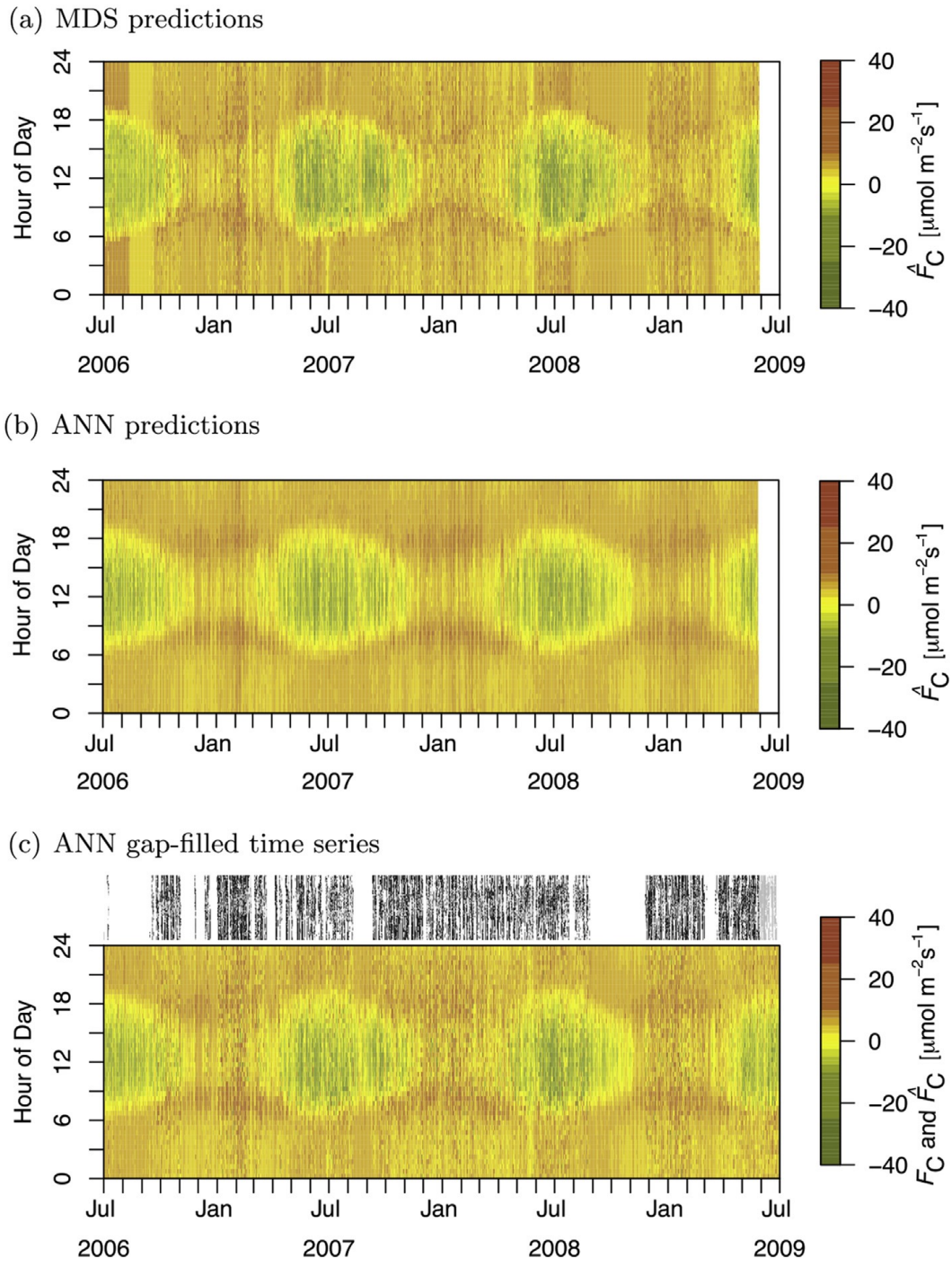
- Traffic is an important driving variable for gap-filling models at a suburban site.
- The authors are not aware of previous attempts to estimate systematic uncertainties at urban sites by the methods in this study.

Significance:

- Urban areas account for at least 70% of anthropogenic CO_2 emissions, and provide opportunities for emission reduction through urban design choices.
- A number of flux towers in urban environments currently measure surface atmospheric exchanges of CO_2 by the EC method and the network is expanding.
- EC provides a means to directly measure the net CO_2 flux, but is currently limited by fragmentation of data sets due to system failures (such as snow, lightning or birds) and low turbulence atmospheric conditions (during the night), which can result in rejection of observations (gaps). Gaps typically account for 20-60% of a flux data set on annual basis.
- Currently available gap-filling models have been designed for homogeneous terrain and for sites in which biological processes predominate, rather than complex terrain, such as urban areas.
- Gap-filling approaches that are widely used for measurements from towers in natural vegetation are based on light and temperature response models; they do not account for key features of the urban environment such as tower footprint heterogeneity and localized CO_2 sources.
- Our results suggest that gap-filling for urban eddy covariance sites may be improved using new models that explicitly incorporate wind direction.
- The gap-filling approach developed here may also be useful for sites with complex terrain other than urban areas, such as logged forests or ecosystems under disturbance from fire or pests.
- The machine learning regression methods are not limited to the variables used in this study, but rather are flexible enough to incorporate other site-specific variables, such as data from footprint models, or information about anthropogenic emission sources other than traffic.



The gap-filling modeling framework depicted as a flowchart. Quality filtered data were first randomly re-sampled, then sets of six driving variables were randomly assigned per model. Three different methods (Artificial Neural Network e ANN, Radial Basis Function Network e RBF, Gaussian Process e GP) were subsequently used for model training before the best 10 out of 1000 models for each wind sector were selected and used for prediction.



Annual course of MDS (a) and ANN (b) bFC model predictions and the observed FC time series, gap-filled with ANN predictions, for the entire time span (c). Data quality is shown above plot (c), with black signifying high quality data, gray for measurements that were of lower quality and not used for model training, and white indicating missing data points. Note the ANN's greater ability to reproduce the diurnal cycle in regions of missing data. We filled short gaps (<2 h) by linear interpolation and there were 552 gaps in May and June 2009 due to missing meteorological data that were filled by mean diurnal cycles for this month. All other gaps were filled by the wind direction dependent gap-filling model presented here.